JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 65, 638–659 (1978)

On Some Aspects in Stochastic Dynamic Programming with Terminal Region

HIDEKI KOKAME

Institute of Atomic Energy, Kyoto University, Uji, Kyoto, Japan

Submitted by M. Aoki

1. INTRODUCTION

A stochastic control problem with terminal region is considered making use of Markov controlled model. The familiar framework of the problem is as follows. Let the terminating time τ of control process be defined as the first entrance time into the terminal region T and required to be finite with probability 1. The expectation of a cost functional which evaluates the system behavior until τ, is usually minimized with respect to the admissible control law somehow defined. Since the terminating time is of stochastic nature, the problem may be considered to have infinite horizon.

Roughly speaking, this paper is concerned with two subjects associated with the approach of stochastic dynamic programming. Supposing that the state space does not include any topology, one of them is to provide a sufficient condition for measurability of the optimum cost function, which will be defined pointwise, and to show that under the same condition optimality equation is fulfilled everywhere in the state space.

Another subject is originated from the question about computability of the optimum cost function and existence of optimal control law. This question relates closely to the uniqueness property of the optimality equation. In connection with this point, let the control problem be generalized so that the control process may be permitted to continue infinitely, while a newly introduced cost $\chi(x_1, x_2, \ldots)$ is imposed on this infiniteness. Of course the problem reduces to the original one when $\chi = \infty$. Dynamic programming approach then yields the same necessary condition for the optimum cost functions g_χ under a natural assumption for χ, which means that the optimality equation need not have a unique solution. The circumstances may well illustrate the difficulty of finding the optimum cost function desired. In accordance with the state of affairs, the optimum cost functions g_χ are studied as a generalized version of the problem.

The organization of the paper is as follows. Section 2 prepares the mathematical preliminaries that are necessary in order to consider the infinite horizon case, and other descriptions of materials. The former subject is discussed in

0022-247X/78/0653-0638$02.00/0

Section 3, and subsequent sections are devoted to the latter one. Section 4 fully discloses the characteristics of the optimum cost functions g_χ with the aid of martingale theory. Among all χ, the cases $\chi = \infty$ and 0 seem to be most serviceable in view of application. In Section 5, these cases are examined in detail and a heuristic algorithm is presented for g_∞. Moreover, a problem of practical interest is analyzed as an example possessing another χ. Section 6 contains the conditions under which the optimality equation has some uniqueness property, for the most desirable case. Even if the optimum cost function has been obtained, optimal or suboptimal control does not always follow directly. Section 7 discusses on this point, though insufficient, including a composite method for suboptimal control in the case of $\chi = \infty$.

2. Preliminaries

Notations are illustrated in the first place. The state space discussed here is assumed to be a measurable space $(X, \mathfrak{F})$. The n-fold and countably infinite direct products of $(X, \mathfrak{F})$'s are represented by $(X^n, \mathfrak{F}_n)$, $(X^\infty, \mathfrak{F}_\infty)$ respectively, and an element of X^∞ is denoted by $\mathbf{x} = (x_1, x_2, \ldots)$. A terminal region $T \in \mathfrak{F}$ is provided on the state space. The complement of the set T, called continuation region, is written by C. The trace of σ-algebra $\mathfrak{F}$ on C is denoted by $\mathfrak{F}^C$, and those direct products of $(C, \mathfrak{F}^C)$'s are represented by $(C^n, \mathfrak{F}_n{}^C)$, $(C^\infty, \mathfrak{F}_\infty{}^C)$ correspondingly. The control space U is assumed to be a compact metric space, whose Borel field is written by $\mathfrak{B}$. Since the *perfectly observable case* is considered in this paper, we can take the control law ρ as the set of functions $\{u_i(x_1, \ldots, x_i);\ i = 1, 2, \ldots\}$ without loss of generality. However, $u_i(x_1, \ldots, x_i)$ may be a $\mathfrak{F}_i{}^C$-measurable function defined on C^i, i.e.,

$$\{(x_1, \ldots, x_i) \mid u_i(x_1, \ldots, x_i) \in B\} \in \mathfrak{F}_i{}^C$$

for an arbitrary $B \in \mathfrak{B}$ and satisfy

$$u_i(x_1, \ldots, x_i) \in U(x_i)\ (\subset U),$$

where $U(x)$, $x \in C$, is a provided state-dependent control region. The set of all control laws that satisfy these conditions is denoted by R.

For the generalized problem stated in Introduction, total cost under the control law ρ can be expressed by a $\mathfrak{F}_\infty$-measurable function

$$z(\rho, \chi) = \sum_{i=1}^{\tau-1} q(x_i, u_i(x_1, \ldots, x_i)) + I_{\{\tau<\infty\}} r(x_\tau) + I_{\{\tau=\infty\}} \chi(x_1, x_2, \ldots),$$

where $q(x, u)$ denotes the continuation cost and $r(x)$ the terminal cost.

In order to regard the expected total cost as a function of the initial state x, A probability $P_x{}^\rho$ on $(X^\infty, \mathfrak{F}_\infty)$ has to be uniquely determined for any $\rho \in R$. Let the transition probability $P(x, u; A)$, $x \in C$ and $A \in \mathfrak{F}$, be given including the control variable u. When the sequence of states $(x_1, ..., x_n) \in C^n$ is observed at time n, the control law ρ defines a probability of x_{n+1} according to

$$P_n{}^\rho(x_1, ..., x_n; A) = P(x_n, u_n(x_1, ..., x_n); A), \qquad A \in \mathfrak{F}.$$

Suppose that

$$P_n{}^\rho(x_1, ..., x_n; A) = I_A(x_k)$$

for $(x_1, ..., x_n) \notin C^n$, where $I_A(\cdot)$ denotes the characteristic function of the set A and k is the first i such that $x_i \in T$. As is shown with ease, the assumption (A1)–1 in next section assures that $P_n{}^\rho$ is a $\mathfrak{F}_n$-measurable function for every $A \in \mathfrak{F}$, and a probability on $(X, \mathfrak{F})$ for every $(x_1, ..., x_n) \in X^n$. Then, the desired probability $P^\rho_{x_1 \cdots x_n}$ on $(X^\infty, \mathfrak{F}_\infty)$ that satisfies

$$\begin{aligned} P^\rho_{x_1,...,x_n}\left[\prod_{i=1}^{k} F_i\right] = I_{F_1}(x_1) \cdots I_{F_n}(x_n) \int_{F_{n+1}} P_n{}^\rho(x_1, ..., x_n; dx_{n+1}) \\ \cdots \int_{F_k} P_k{}^\rho(x_1, ..., x_{k-1}; dx_k) \end{aligned} \tag{1}$$

for an arbitrary rectangular set $\prod_{i=1}^k F_i \in \mathfrak{F}_k$, is uniquely determined by Tulcea's theorem (e.g., Neveu [1]). Moreover, if an initial probability η on $(X^1, \mathfrak{F}_1)$ is given,

$$P_\eta{}^\rho(A) = \int \eta(dx_1)\, P^\rho_{x_1}(A), \qquad A \in \mathfrak{F}_\infty,$$

defines a probability on $(X^\infty, \mathfrak{F}_\infty)$.

Now letting $E_x{}^\rho$ denote the expectation with respect to $P_x{}^\rho$, the function $E_x{}^\rho z^\chi$ is called cost function for ρ, which is measurable in x. Then the optimum cost function is defined pointwise by the infimum of the cost functions:

$$g_\chi(x) \triangleq \inf_{\rho \in R} E_x{}^\rho z^\chi. \tag{2}$$

From now on, the expressions $z(\rho, \chi)$ and $q(x_i, u_i(x_1, ..., x_i))$ will be mostly abbreviated to z^χ and $q(x_i, u_i)$ respectively, when the control law associated is clear from the context. The law whose cost function is less than $g_\chi(x) + \epsilon$ uniformly is called ϵ-optimal, and 0-optimal is simply termed optimal.

For each control problem with an initial probability η, our interest is to find $\inf_{\rho \in R} E_\eta{}^\rho z^\chi$ and obtain the control law that attains the infimum if exists. The law whose expected cost is less than the infimum plus ϵ, is called (η, ϵ)-optimal.

3. Optimality Equation and Measurability of the Optimum Cost Function

For stochastic control problems with infinite horizon, some authors have worked on the measurability of the optimum cost function. The function was proved by Blackwell [2] to be a Baire function if an ϵ-optimal control exists for $\epsilon > 0$. Under a general framework, Strauch [3] showed that it is universally measurable. Hiderer [4] extended this result to the nonstationary case, and also obtained some theorems for the case that the state space is denoted by a measurable space. This section presents another result on the measurability under a simple assumption (A1)–2 below, deriving the optimality equation in an expository style via a strict version of dynamic programming approach [5].

The following assumptions are effective throughout this paper.

(A1) The transition probability $P(x, u; A)$, $x \in C$ and $A \in \mathfrak{F}$, satisfies

(1) $P(x, u; \cdot)$ is a probability measure on $(X, \mathfrak{F})$, $P(\cdot, u; A)$ is measurable on C, and $P(x, \cdot; A)$ is continuous,

(2) there exists a measure μ on $(C, \mathfrak{F}^C)$ such that $P(x, u; \cdot)$ is absolutely continuous with respect to μ, if considered as a finite measure on $(C, \mathfrak{F}^C)$.

(A2) The nonnegative functions $r(x)$ and $q(x, u)$ are everywhere finite, $r(\cdot)$ and $q(\cdot, u)$ are measurable on T and C respectively, and $q(x, \cdot)$ is continuous.

(A3) The nonnegative $\mathfrak{F}_\infty$-measurable function $\chi(\mathbf{x})$ satisfies

$$\chi(x_1, \ldots, x_n, x_{n+1}, \ldots) = \chi(x_{n+1}, x_{n+2}, \ldots)$$

for every $\mathbf{x}$ and n.

(A4) The control region $U(x)$ is a nonempty compact subset in U and satisfies

$$\{x \mid U(x) \cap F \neq \varnothing\} \in \mathfrak{F}^C$$

for an arbitrary closed subset $F \subset U$.

(A5) The set of control laws R_x is not empty, and there exists a measurable function $\xi(x)$ such that

$$\int P(x, u; dy)\, \xi(y) < \infty$$

and

$$E_x^{\hat{\rho}} z^{\chi} \leqslant \xi(x)$$

for some $\hat{\rho} \in R_x$, where R_x is defined by

$$\{\rho \mid \rho \in R \text{ and } E_x^{\rho} z^{\chi} < \infty \text{ for all } x\}.$$

For the convenience of description we shall make use of some notations about control laws. Given a control law $\rho \triangleq \{u_j(x_1,\ldots,x_j); j = 1, 2,\ldots\}$ and measurable functions $u_j'(x_1,\ldots,x_j) \in U(x_j)$, $1 \leqslant j \leqslant n$, the functions

$$\begin{aligned} \bar{u}_j(x_1,\ldots,x_j) &= u_j'(x_1,\ldots,x_j), & 1 \leqslant j \leqslant n, \\ &= u_{j-n}(x_{n+1},\ldots,x_j), & j > n, \end{aligned}$$

define a control law clearly. Let this control law be denoted by $\{u_j'(x_1,\ldots,x_j); 1 \leqslant j \leqslant n\} \oplus \rho$. For a control law ρ, and $(\bar{x}_1,\ldots,\bar{x}_n)$ such that $\tau(\bar{x}_1,\ldots,\bar{x}_n) > n$, the functions $u_{n+i}(\bar{x}_1,\ldots,\bar{x}_n, x_1,\ldots,x_i)$; $i = 1, 2,\ldots$, also define a control law, which is expressed by $\rho(\bar{x}_1,\ldots,\bar{x}_n)$. Then the equation

$$E^{\rho}_{\bar{x}_1\cdots\bar{x}_{n+1}} z^X = \sum_{i=1}^{n} q(\bar{x}_i, u_i(\bar{x}_1,\ldots,\bar{x}_i)) + E^{\rho(\bar{x}_1,\ldots,\bar{x}_n)}_{\bar{x}_{n+1}} z^X \tag{3}$$

is justified on account of (1) and the assumption (A3), if $\tau(\bar{x}_1,\ldots,\bar{x}_n) > n$. Utilizing these notations, the relation

$$[\{u_j'; 1 \leqslant j \leqslant n\} \oplus \rho]\,(\bar{x}_1,\ldots,\bar{x}_n) = \rho \tag{4}$$

holds for every $(\bar{x}_1,\ldots,\bar{x}_n)$ such that $\tau(\bar{x}_1,\ldots,\bar{x}_n) > n$.

Also note that for given control laws ρ^1 and ρ^2 $(\in R)$, there exists a control law ρ^0 such that

$$E_x^{\rho^0} z^X = \min_{i=1,2} E_x^{\rho^i} z^X. \tag{5}$$

In fact, the control law ρ^0 is such one that takes ρ^1 if the initial state x satisfies $E_x^{\rho^1} z^X \leqslant E_x^{\rho^2} z^X$, and otherwise takes ρ^2.

THEOREM 1. *The optimum cost function $g_X(x)$ is measurable and satisfies the optimality equation*

$$\begin{aligned} g_X(x) &= r(x), & x \in T, \\ &= \min_{u \in U(x)} \left[q(x, u) + \int P(x, u; dy)\, g_X(y) \right], & x \in C, \end{aligned} \tag{6}$$

everywhere. Furthermore for an arbitrary initial probability η,

$$\inf_{\rho \in R} E_\eta{}^\rho z^X = \int g_X(x)\, \eta(dx). \tag{7}$$

Proof. Let us introduce a measurable function

$$\bar{g}_X(x) = \operatorname*{ess\,inf}_{\rho \in R} E_x{}^\rho z^X, \qquad x \in C,$$

where the essential infimum is taken with respect to the measure μ assumed in (A1)–2. We shall assume $\bar{g}_X(x) = r(x)$ on T, in keeping with the self-evident result that $g_X(x) = r(x)$ for every $x \in T$. On the essential infimum, we note as a preliminary that there exists a sequence $\rho^i \in R$ such that

$$\bar{g}_X(x) = \inf_{i=1,2,\ldots} E_x^{\rho^i} z^X \qquad \mu\text{-a.s.} \quad \text{on} \quad C$$

(see [1], pp. 44–45). Then applying the procedure below (5) to these ρ^i and $\hat{\rho}$ in (A5), we can reconstruct the sequence $\rho^i \in R$ by induction so that their cost functions $\varphi_i(x) = E_x^{\rho^i} z^X$ satisfy $\varphi_i(x) \leqslant \xi(x)$, and converge to $\bar{g}_X(x)$ nonincreasingly;

$$\lim_{i\to\infty} \varphi_i(x) = \bar{g}_X(x) \qquad \mu\text{-a.s.} \quad \text{on} \quad C.$$

To begin with we shall prove (12) seen in the sequal, instead of (6). To show the inequality "$\leqslant$" for (12), control laws $\hat{\rho}^i = \{u(x_1)\} \oplus \rho^i$ are introduced, where $u(x)$ ($\epsilon\ U(x)$) is a measurable function fixed arbitrarily. By (3) and (4), we obtain

$$\begin{aligned} E_{x_1 x_2}^{\hat{\rho}^i} z^X &= q(x_1, u(x_1)) + E_{x_2}^{\hat{\rho}^i(x_1)} z^X \\ &= q(x_1, u(x_1)) + E_{x_2}^{\rho^i} z^X, \qquad x_1 \in C. \end{aligned} \tag{8}$$

Taking expectation of (8) with respect to $P_{x_1}^{\hat{\rho}^i}$, we obtain

$$E_{x_1}^{\hat{\rho}^i} z^X = q(x_1, u(x_1)) + \int P(x_1, u(x_1); dx_2)\, \varphi_i(x_2).$$

Then applying the monotone convergence theorem to the right-hand side, and noting that the definition of $g_X(x)$ implies $g_X(x_1) \leqslant E_{x_1}^{\hat{\rho}^i} z^X$, one may deduce the inequality $g_X(x_1) \leqslant \psi(x_1, u(x_1); \bar{g}_X)$ everywhere on C, where

$$\psi(x, u; s) \triangleq q(x, u) + \int P(x, u; dy)\, s(y).$$

For a *simple function* $s(x)$, the function ψ is continuous in u and measurable in x due to the assumptions (A1)–1 and (A2). Hence, choosing simple functions $\bar{g}_k(x)$ such that $\bar{g}_k(x) \uparrow \bar{g}_X(x)$ as $k \to \infty$, $\psi(x, u; \bar{g}_X)$ is shown to satisfy the conditions appeared in Theorem 8 of Filippov's type whose result need not topology of the state space (see Appendix). Consequently, a measurable function $u^0(x)$ exists that minimizes $\psi(x, u; \bar{g}_X)$ for each $x \in C$, where u ranges over $U(x)$. Then it follows that

$$g_X(x) \leqslant \min_{u \in U(x)} \psi(x, u; \bar{g}_X), \qquad x \in C. \tag{9}$$

Conversely for the inequality "$\geqslant$", remark that

$$E^{\rho}_{x_1x_2}z^X = q(x_1, u_1(x_1)) + E^{\rho(x_1)}_{x_2}z^X \tag{10}$$

for every $\rho \in R$ and $x_1 \in C$. From $\rho(x_1) \in R$, the inequality $E^{\rho(x_1)}_{x}z^X \geqslant \bar{g}_x(x)$ is established μ-a.s. on C. With this inequality and the assumption (A1)–2, the expectation of (10) with respect to $P^{\rho}_{x_1}$ reduces to

$$E^{\rho}_{x_1}z^X \geqslant \psi(x_1, u_1(x_1); \bar{g}_x) \geqslant \min_{u \in U(x_1)} \psi(x_1, u; \bar{g}_x), \tag{11}$$

for each $x_1 \in C$. Recalling that $\rho \in R$ is arbitrary, the relevant inequality is proved. Thus we obtain:

$$g_x(x) = \min_{u \in U(x)} \left[q(x, u) + \int P(x, u; dy)\, \bar{g}_x(y)\right]. \tag{12}$$

As a consequence of (12), the measurability of $g_x(x)$ is confirmed by Lemma 5 in Appendix.

To complete the proof of (6), it suffices to show $g_x(x) = \bar{g}_x(x)$ μ-a.s. on account of (12) and (A1)–2. Note that the latter representation of $\bar{g}_x$ implies $g_x(x) \leqslant \bar{g}_x(s)$ μ-a.s. For the reverse inequality, suppose that $u^0(x)$ described above be substituted in (8). Then, by (12), the sequence $\bar{\rho}^i = \{u^0(x)\} \oplus \rho^i$ satisfies

$$E^{\bar{\rho}^i}_{x_1}z^X \downarrow \psi(x_1, u^0(x_1); \bar{g}_x) = g_x(x_1), \qquad x_1 \in C, \tag{13}$$

as i tends to infinity. This verifies the desired inequality.

To prove (7) finally, let us define measurable sets $A_i \in \mathfrak{F}^C$ for a given $\epsilon > 0$;

$$A_i = \{x \mid E^{\bar{\rho}^i}_{x}\, z^X \leqslant g_x(x) + \epsilon\} \Big/ \bigcup_{k=1}^{i-1} A_k, \qquad i = 1, 2, \ldots.$$

The relation (13) then makes the sets A_i a division of C. Thus, expressing $\bar{\rho}^i$ as $\{\bar{u}_j{}^i; j = 1, 2, \ldots\}$, the control law ρ^ϵ that has the control functions

$$u_j{}^\epsilon(x_1, \ldots, x_j) = \sum_{i=1}^{\infty} I_{A_i}\bar{u}_j{}^i(x_1, \ldots, x_j),$$

satisfies $E^{\rho^\epsilon}_{x_1}z^X \leqslant g_x(x_1) + \epsilon$ for every $x_1 \in C$. Hence the cost functions of ρ^ϵ converge to $g_x(x_1)$ uniformly as ϵ tends to 0. Therefore, the equation (7) holds including the possibility that both sides of (7) are infinity.

Note that the control law ρ^ϵ in the proof is ϵ-optimal. Theorem 1 can thus be considered to give a simple case where the conditions in Theorem 6 of [2] or Theorem 19.3 of [4] hold, while both of them were proved for the *discounted case.*

Remark. The assumption (A1) has to be checked at first in applying these results to stochastic control systems usually described by

$$x_{i+1} = f(x_i, u_i, w_i), \qquad i = 1, 2, \ldots,$$

where x_i, u_i, and w_i represent d_s-state vector, d_c-control vector, and input noise of d_n-dimension at time i, respectively. Moreover the sequence $\{w_i\}$ is assumed to be a white stationary process. For this case, the transition probability may be formally defined by

$$P(x, u; A) = \int I_A(f(x, u, w))\, dW(w),$$

where W denotes the probability which w_i obeys, and A is a Borel set in d_s-Euclidean space. However it is rather difficult to provide a general structure of the system that satisfies (A1). Let us consider the special example

$$f(x, u, w) = f_1(x, u) + f_2(x)\, w.$$

Then the assumption (A1) is established under the following conditions:

1) $f_1(x, u)$ is Borel measurable in x and continuous in u,

2) $f_2(x)$ is a Borel measurable $d_s \times d_n$ matrix whose rank is d_s,

3) the probability distribution W has a bounded density function $\theta(w)$ with respect to Lebesgue measure $\tilde{\mu}$ on d_n-Euclidean space.

In fact, one may exhibit the continuity in (A1)–1, referring to the relation

$$\lim_{\epsilon \to 0} \int | \varphi(w + \epsilon) - \varphi(w) |\, \tilde{\mu}(dw) = 0$$

for a positive function φ. And the assumption (A1)–2 can be easily obtained by the conditions 2) and 3), taking Lebesgue measure $\tilde{\mu}$ as μ.

In closing this section, note that the assumption (A5) enables the definition (1) to be replaced by

$$g_\chi(x) = \inf_{\rho \in R_\chi} E_x{}^\rho z^\chi, \tag{14}$$

on account of the property (5).

4. Characterization of the Optimum Cost Function

The optimum cost functions g_χ are to differ intrinsically from each other since the criterion z^χ varies with the cost χ. Theorem 1 however shows that these functions satisfy the same nonlinear equation (6). Thus, an auxiary characterization of g_χ's is necessary in order to identify the optimum cost function desired from among the solutions of (6). This subject at issue is completely solved in Theorem 2 below.

Theorem 2 is based on the attribute that the following sequence $z(n; \varphi)$ makes a submartingale under any control law $\rho \in R$ if φ is a solution of (6):

$$z(n; \varphi) = \sum_{i=1}^{\tau(n)-1} q(x_i, u_i) + \varphi(x_{\tau(n)}), \qquad n = 1, 2,..., \tag{15}$$

where $\tau(n) = \min(\tau, n)$. In fact, using the equality

$$E^{\rho}_{x_1 \cdots x_n} \varphi(x_{n+1}) = \int P(x_n, u_n; dy)\, \varphi(y) \qquad \text{on} \qquad \{\tau > n\},$$

we obtain:

$$\begin{aligned} E^{\rho}_{x_1 \cdots x_n} z(n+1; \varphi) &= I_{\tau \leqslant n} z(n; \varphi) + I_{\tau > n} \left[\sum_{i=1}^{n} q(x_i, u_i) + \int P(x_n, u_n; dy)\, \varphi(y) \right] \\ &\geqslant I_{\tau \leqslant n} z(n; \varphi) + I_{\tau > n} \left[\sum_{i=1}^{n-1} q(x_i, u_i) + \varphi(x_n) \right] \\ &= z(n; \varphi). \end{aligned}$$

THEOREM 2. *Let Φ_χ denote the family of the solutions of* (6) *such that for all x and $\rho \in R_\chi$,*

1) *the sequence $z(n; \varphi)$ is uniformly integrable with respect to $P_x{}^\rho$,*
2) $\chi(x_1, x_2, ...) \geqslant \liminf_{n \to \infty} \varphi(x_n)$ $P_x{}^\rho$-a.s. *on* $\{\tau = \infty\}$.

Then the optimum cost function $g_\chi(x)$ belongs to Φ_χ, and the inequality $g_\chi(x) \geqslant \varphi(x)$ holds for an arbitrary $\varphi \in \Phi_\chi$. Namely, g_χ is the maximum element of the family.

Proof. First the inclusion $g_\chi \in \Phi_\chi$ is proved. For the condition 1), we shall show the inequality

$$E^{\rho}_{x_1 \cdots x_n} z^\chi \geqslant z(n; g_\chi) \tag{16}$$

for each $(x_1, ..., x_n) \in X^n$, letting a control law $\rho \in R_\chi$ be fixed arbitrarily. Obviously,

$$E^{\rho}_{x_1 \cdots x_n} z^\chi = I_{\tau < n} z^\chi + I_{\tau > n-1} E^{\rho}_{x_1 \cdots x_n} z^\chi. \tag{17}$$

By (15), noting the relation $g_\chi(x_\tau) = r(x_\tau)$,

$$z(n; g_\chi) = I_{\tau < n} z^\chi + I_{\tau > n-1} \left[\sum_{i=1}^{n-1} q(x_i, u_i) + g_\chi(x_n) \right]. \tag{18}$$

Since the definition of g_χ allows

$$g_\chi(x_n) \leqslant E_{x_n}^{\rho(x_1,\ldots,x_{n-1})} z^\chi \qquad \text{on} \qquad \{\tau > n-1\},$$

equations (17) and (18) yield (16) with the aid of (3).

It then suffices to show the uniform integrability of $E^\rho_{x_1,\ldots,x_n} z^\chi$. Note that this sequence of $\mathfrak{F}_n$-measurable functions forms a martingale with respect to $P_x{}^\rho$. Obviously, this sequence remains a martingale even if $\mathfrak{F}_\infty$-measurable function z^χ is associated as the case $n = \infty$. Hence, on account of the martingale convergence theorem (e.g., Doob [6]), we can obtain the uniform integrability desired.

The condition 2) is shown by taking the limit of (16). In fact, the left-hand side of (16) converges $P_x{}^\rho$-a.s. to z^χ as n approaches infinity. On the other hand, since g_χ is a solution of (6), the sequence $z(n; g_\chi)$ makes a martingale as previously noted. Moreover we have

$$\sup_n E_x{}^\rho z(n; g_\chi) \leqslant E_x{}^\rho z^\chi < \infty,$$

due to (16) and $\rho \in R_\chi$. The submartingale convergence theorem then applies, and the sequence $z(n; g_\chi)$ is proved to converge $P_x{}^\rho$-a.s. to

$$I_{\tau<\infty} z^\chi + I_{\tau=\infty}\left[\sum_{i=1}^{\infty} q(x_i, u_i) + \lim_{n\to\infty} g_\chi(x_n)\right].$$

Comparing the limit functions, the condition 2) is approved in a stronger form.

For the maximum property of g_χ, let us take an arbitrary function $\varphi \in \Phi_\chi$. The remark preceding Theorem 2 implies the inequality $\varphi(x) \leqslant E_x{}^\rho z(n; \varphi)$ for each x and $\rho \in R$. By the condition 1), Fatou's lemma then assures

$$\varphi(x) \leqslant E_x{}^\rho \limsup_{n\to\infty} z(n; \varphi) \tag{19}$$

for $\rho \in R_\chi$. Due to (18) and the inequality in the condition 2), the relation

$$\limsup_{n\to\infty} z(n; \varphi) \leqslant z^\chi, \qquad P_x{}^\rho\text{-a.s.}, \tag{20}$$

is derived for $\rho \in R_\chi$. Substituting (20) into (19), we obtain $\varphi(x) \leqslant E_x{}^\rho z^\chi$ for an arbitrary control law $\rho \in R_\chi$. Then, the alternative definition of g_χ appeared in (14) yields required inequality $\varphi(x) \leqslant g_\chi(x)$.

For the case $\chi = \infty$, it is to be noted that the condition 1) only remains to be active since the condition 2) always holds for every φ.

5. THE CASES $\chi = \infty$ AND 0

Special interest might be taken in the cases $\chi = \infty$ and 0. In the former case, as noted in Introduction, the optimum cost function $g_\infty(x)$ can be considered as that of the problem where the control law is to be optimized within

$$R^f \triangleq \{\rho \mid P_x{}^\rho(\tau < \infty) = 1 \text{ for all } x\},$$

i.e., the set of control laws that attain the terminal region T w.p.l. starting from an arbitrary x. This is attributable to (14) and the inclusion relation $R_\infty \subset R^f \subset R$. The requirement that the control law be restricted to R^f seems to be met with frequently in practical problems.

In the latter case, the formulation is an obvious extension of negative dynamic programming [4] to the case with terminal region. In this section, discussions will be focussed primarily on these cases.

Since Theorem 2 in itself does not suggest a constructive method for $g_\chi(x)$, it seems very difficult to work out a method of general use. However, some heuristic method can be proposed for the optimum cost function $g_\infty(x)$.

THEOREM 3. *Let a measurable function $\psi(x)$ satisfy the following conditions:*

1) $\psi(x) \geqslant E_x^{\rho^1} z^\infty$ *holds everywhere for some* $\rho^1 \in R_\infty$,

2) *the sequence $z(n; \psi)$ is uniformly integrable with respect to $P_x{}^\rho$ for each $\rho \in R_\infty$ and x,*

3) $\psi(x) \geqslant \min_{u \in U(x)} \left[q(x, u) + \int P(x, u; dy)\, \psi(y)\right]$

holds for $x \in C$. Then the recursion relation

$$\begin{aligned} \psi_{n+1}(x) &= r(x), && x \in T, \\ &= \min_{u \in U(x)} \left[q(x, u) + \int P(x, u; dy)\, \psi_n(y)\right], && x \in C, \end{aligned} \tag{21}$$

with initial function

$$\psi_1(x) = \psi(x), \tag{22}$$

generates a nonincreasing sequence $\psi_n(x)$ that converges to the optimum cost function $g_\infty(x)$.

Proof. Considering $\psi_{n+1}(x)$ as the image of $\psi_n(x)$ in (21), let us define an operator Q. Since the inequality $\psi(x) \geqslant r(x)$ is true on T due to the condition 1), the condition 3) implies $\psi_1 \geqslant Q\psi_1$. Noting that the operator has monotonicity, i.e., if $\varphi_1 \geqslant \varphi_2$ then $Q\varphi_1 \geqslant Q\varphi_2$ holds, the inequality $Q^{n+1}\psi_1 \leqslant Q^n\psi_1$ is derived, which means that the sequence $\psi_n(x)$ is nonincreasing.

Letting $\psi_\infty(x)$ denote the limit of this sequence, $\psi_\infty(x)$ is easily shown to satisfy (6). Then, the condition 2) assures the inclusion $\psi_\infty \in \Phi_\infty$ as was noted at the end of Section 4. Thus $\psi_\infty(x) \leqslant g_\infty(x)$ is obtained.

Now the proof is completed if the reverse inequality is shown. Let $\lambda_n(x)$ be the function obtained by $(n-1)$-times recursive substitutions into (21) with initial function $E_x^{\rho^1} z^\infty$. The monotonicity of Q yields

$$\psi_n(x) \geqslant \lambda_n(x), \tag{23}$$

on account of the condition 1). Now consider the equation (21) for $\lambda_2(x)$. Then there exists a measurable function $u^2(x) \in U(x)$ that attains the minimum of the right-hand side, similarly as was noted for $g_\chi(x)$ in Section 3. Taking the control law $\rho^2 \triangleq \{u^2(x)\} \oplus \rho^1$, we obtain $g_\infty(x) \leqslant E_x^{\rho^2} z^\infty = \lambda_2(x)$ from the definition of g_∞. The relation $g_\infty(x) \leqslant \lambda_n(x)$ is proved inductively. Consequently, we obtain $g_\infty(x) \leqslant \psi_n(x)$ from (23), which verifies the required inequality.

Theorem 3 is crucially based on the existence of $\psi(x)$. Fortunately we are confirmed of its existence. In fact, the function $\psi(x) = g_\infty(x) + \epsilon$ ($\epsilon > 0$) is the relevant, since the control law ρ^ϵ in the proof of Theorem 1 satisfies condition 1). The other conditions are easily checked.

Difficulties in applying this theorem are mainly ascribed to the condition 2). A useful sufficient condition is contained in the following lemma. Note $R_\infty \subset R_0$.

LEMMA 1. *If $\psi(x) = r(x)$ on T, and there exist positive constants M_1, M_2 such that*

$$\psi(x) \leqslant M_1 q(x, u) + M_2, \tag{24}$$

holds for all $u \in U(x)$ and $x \in C$, then the sequence $z(n; \psi)$ is uniformly integrable with respect to P_x^ρ for any $\rho \in R_0$.

Proof. Invoking the equation (15) written as

$$z(n; \psi) = \sum_{i=1}^{\tau(n)-1} q(x_i, u_i) + I_{\tau \leqslant n} r(x_\tau) + I_{\tau > n} \psi(x_n),$$

note that the first two terms of the right-hand side are dominated by the total cost z^0. Then the terms are uniformly integrable with respect to P_x^ρ if $\rho \in R_0$, since $E_x^\rho z^0 < \infty$ holds. As for the last term, with use made of (24), we obtain

$$I_{\tau > n} \psi(x_n) \leqslant M_1 I_{\tau > n} q(x_n, u_n) + M_2 \leqslant M_1 \sum_{i=1}^{\tau-1} q(x_i, u_i) + M_2.$$

This inequality leads to the conclusion in a similar way.

Next, the case $\chi = 0$ is dealt with. Theorem 4 informs that this case is the simplest among all χ; stationary optimal control exists and is obtained directly from the optimum cost function, where a control law is called *stationary* if its control functions are written as $u_i(x_1, ..., x_i) = u_1(x_i)$ for some measurable $u_1(x) \in U(x)$.

Existence of stationary optimal control has been investigated by several authors. Maitra [7] first presented a sufficient condition for the discounted case. The state space being a Borel set, extensions to the state-dependent control region were given by Furukawa [8] and Schäl [9]. Now, Theorem 4 provides a simple result for negative dynamic programming on a measurable state space.

THEOREM 4. *The optimum cost function $g_0(x)$ is the minimal solution of* (6). *Moreover, the control law $\rho^0 \triangleq \{u^0(x_i); i = 1, 2,...\}$ is optimal, i.e.,*

$$E_x^{\rho^0} z^0 = g_0(x),$$

where $u^0(x)$ is such that for all $x \in C$,

$$g_0(x) = q(x, u^0(x)) + \int P(x, u^0(x); dy)\, g_0(y).$$

Proof. Let $\varphi(x)$ be the minimal solution. Then clearly follows $\varphi(x) \leqslant g_0(x)$. To show the reverse inequality, let control law $\bar{\rho}$ be constructed from $\varphi(x)$ as ρ^0 from $g_0(x)$. Note that the sequence $z(n; \varphi)$ is a martingale under $\bar{\rho}$;

$$\varphi(x) = E_x{}^{\bar{\rho}} \left[\sum_{i=1}^{\tau(n)-1} q(x_i, u_i) + I_{\tau \leqslant n} r(x_\tau) + I_{\tau > n} \varphi(x_n) \right].$$

On account of Fatou's lemma, we can derive

$$\begin{aligned} \varphi(x) &\geqslant E_x{}^{\bar{\rho}} \left[I_{\tau<\infty} z^0 + I_{\tau=\infty} \left\{ \sum_{i=1}^{\infty} q(x_i, u_i) + \liminf_{n\to\infty} \varphi(x_n) \right\} \right] \\ &\geqslant E_x{}^{\bar{\rho}} \left[I_{\tau<\infty} z^0 + I_{\tau=\infty} \sum_{i=1}^{\infty} q(x_i, u_i) \right] = E_x{}^{\bar{\rho}} z^0. \end{aligned} \tag{25}$$

Then, by the definition of g_0, $\varphi(x) \geqslant g_0(x)$ is reduced establishing the minimum property. Optimality of ρ^0 also results from (25), substituting g_0 into φ.

It is to be noticed that we can obtain the minimal solution of (6) from the recursion relation (21), taking 0 as the initial function of iteration. Hereafter, such a control law as ρ^0 is called, for example, corresponding one to the solution $g_0(x)$.

The cases $\chi = \infty$ and 0 have been studied. However, the problems with another χ seem more difficult to solve. Two examples of interest are examined in order to infer the circumstances.

EXAMPLES. Consider a pursuit problem where the object to be maximized is the probability that the state reaches the target set T in finite time. This can be reformulated as the problem of minimizing P_x^{ρ} $(\tau = \infty)$, which is expressed by $r = 0$, $q = 0$, and $\chi = 1$. The optimality equation is

$$\begin{aligned} g_1(x) &= 0, && x \in T, \\ &= \min_{u \in U(x)} \int P(x, u; dy)\, g_1(y), && x \in C. \end{aligned}$$

Theoretical interest may be aroused because this equation always has solutions $ag_1(x)$ $(a \geqslant 0)$.

With $r(x)$ being extended onto C as $r(x) = 1$, the inequality $g_1(x) \leqslant r(x)$ is immediate. Note that any solution φ satisfies the conditions of Theorem 2 if $\varphi(x) \leqslant r(x)$. Indeed the boundedness of $z(n; \varphi)$ and the inequality $\limsup_{n\to\infty} \varphi(x_n) \leqslant 1$ validate the conditions 1) and 2), respectively. Hence $g_1(x)$ is the maximum solution that satisfies $\varphi(x) \leqslant r(x)$, and therefore the recursive substitution into the optimality equation leads to $g_1(x)$ when $r(x)$ is selected for the initial function.

With respect to the optimal control law, both its existence and approachability are however open to question. By appealing to a concrete version, we demonstrate the difficulties inherent to the study of the pursuit problem. Assume that the state space consists of denumerable points 0, 1, 2,..., and the target set T be $\{0\}$. For the state k $(\geqslant 1)$, the transition probability is described as follows:

$$\begin{aligned} P(k, u; m) &= a_k u, && m = k + 1, \\ &= (1 - a_k u)/k, && m = 1, \\ &= (1 - a_k u)(1 - 1/k), && m = 0. \end{aligned}$$

Figure 1 illustrates the state transition with its probability. The control region $U(k)$ is the unit interval [0, 1].

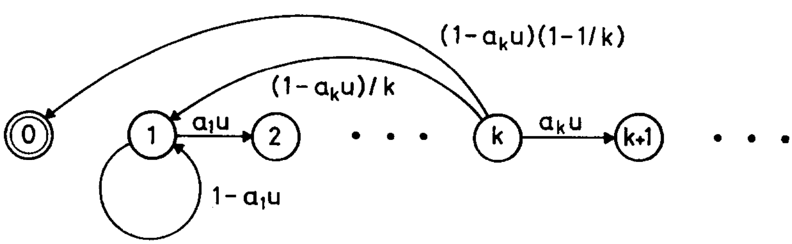

FIG. 1. Conceptual diagram of the state transition.

The first difficulty is that the control law $\rho^0 \triangleq \{u^0(x_i); i = 1, 2, \ldots\}$ corresponding to $g_1(x)$ may not be optimal. To see this, let $a_k = 1 - 1/k^2$. It follows at once that $g_1(1) = 1$, and then

$$g_1(k) = \min_{0 \leqslant u \leqslant 1} [1/k - (1 - 1/k^2)(1/k - g_1(k+1))\, u], \qquad k \geqslant 2.$$

From this equation, we can easily derive $g_1(k) \leqslant 1/k$ for every $k \geqslant 1$, and $u^0(k) = 1$.

Generally, since g_1 is bounded, $z(n; g_1)$ is a uniformly integrable martingale under the corresponding control $\rho^0 = \{1, 1, \ldots\}$:

$$g_1(x) = E_x^{\rho^0} \lim_{n \to \infty} z(n; g_1) = E_x^{\rho^0} I_{\tau=\infty} \lim_{n \to \infty} g_1(x_n).$$

On the other hand, every point in the set $\{\tau = \infty\}$ satisfies either (1) $x_n = 1$ for all sufficiently large n, or (2) $x_{n+1} = x_n + 1$ for all $n \geqslant 1$. The limit of g_1 is one in the former, but zero in the latter. These amount to an equation satisfied by ρ^0:

$$P_k^{\rho^0}(\tau = \infty) = g_1(k) + \prod_{i=k}^{\infty} (1 - 1/i^2).$$

The second term of the right-hand side is clearly positive, which means that ρ^0 is not optimal, and subsequently negates the existence of optimal control law.

Another difficulty regarding the approachability to optimal law would naturally arise in cases where $g_1 \equiv 0$. That is, every control law is corresponding to the optimum cost. In such a case, it would be the most reasonable to construct inductively $\rho^i \triangleq \{u^i(x_1), \ldots, u^2(x_{i-1})\} \oplus \rho$ with ρ fixed arbitrarily and $u^i(x)$ corresponding to $\varphi_i(x)$ the $(i-1)$-th solution of iteration, and ultimately to give the limit control $\rho^0 = \{u^0(x_i); i = 1, 2, \ldots\}$, provided $u^i(x) \to u^0(x)$. But the resulting control once again may not be optimal. Assume that $a_k = 1$ in the above example with $m = 2k$ replacing $m = k + 1$. It is easy to see that

$$\varphi_2(k) = 1/k, \qquad \text{and} \qquad u^2(k) = 0;$$
$$\varphi_i(k) = 1/(k \cdot 2^{i-2}), \qquad \text{and} \qquad u^i(k) = 1, \quad \text{for } i > 2.$$

Thus we have $g_1(k) = 0$ for every k. Note that the limit control $\rho^0 = \{1, 1, \ldots\}$ is the worst one: it misses the target set by transferring the state ever in direction to infinity. Optimal is such that takes 1 or 0 according as the state is 1 or not.

To the contrary, consider an evasion problem which treats minimization of P_x^{ρ} $(\tau < \infty)$. Since this problem is expressed by $r = 1$, $q = 0$, and $\chi = 0$, Theorem 4 assures the optimality of the control corresponding to $g_0(x)$. In the case of the concrete version, the above-mentioned law ρ^0 is the optimal independently of the parameters a_k.

Thus it is to be noticed that the pursuit problem involves far more difficulties than the evasion problem.

6. On the Uniqueness of Solution

In the preceding sections discussions have been made mainly on distinguishing aspects of the optimum cost functions g_χ. However, there seem to exist many cases where the optimality equation (6) has a unique solution. The following theorem is concerned with this point.

Theorem 5. *Suppose that ρ^0 is the optimal control law presented in Theorem 4. Let $\varphi(x)$ denote an arbitrary solution of (6). If the sequence $I_{\tau>n}\varphi(x_n)$ is uniformly integrable with respect to $P_x^{\rho^0}$ and*

$$P_x^{\rho^0}(\tau < \infty) = 1, \tag{26}$$

then $\varphi(x) = g_0(x)$ holds.

Proof. Under the optimal ρ^0, the sequences $z(n; \varphi)$ and $z(n; g_0)$ are submartingale and martingale respectively. Namely, the inequalities $\varphi(x) \leqslant E_x^{\rho^0} z(n; \varphi)$ and $g_0(x) = E_x^{\rho^0} z(n; g_0)$ hold. Note that the minimality of $g_0(x)$ implies $\varphi(x) \geqslant g_0(x)$. Then the inequalities yield

$$\varphi(x) - g_0(x) \leqslant E_x^{\rho^0} I_{\tau>n}\{\varphi(x_n) - g_0(x_n)\} \leqslant E_x^{\rho^0} I_{\tau>n}\varphi(x_n). \tag{27}$$

On the other hand, by the condition (26),

$$\lim_{n\to\infty} I_{\tau>n}\varphi(x_n) = 0 \qquad P_x^{\rho^0}\text{-a.s.}$$

Therefore, the uniform integrability of $I_{\tau>n}\varphi(x_n)$ indicates that the last term of (27) converges to zero. Consequently, the required inequality $\varphi(x) \leqslant g_0(x)$ is derived.

Especially, for coincidence of the optimum cost functions $g_\chi(x)$, a simple result is available in the assertion below.

Corollary. *If the condition (26) is valid, then $g_\chi(x) = g_0(x)$ holds for every χ. Namely ρ^0 is optimal for every problem considered.*

Proof. As shown in the proof of Lemma 1, the uniform integrability of $z(n; g_\chi)$ is equivalent to that of $I_{\tau>n} g_\chi(x_n)$ under any $\rho \in R_0$. Also note that $\rho^0 \in R_\chi$ due to (26). Then, by using Theorem 2, it follows that the sequence $I_{\tau>n} g_\chi(x_n)$ is uniformly integrable with respect to $P_x^{\rho^0}$. The conclusion is now ascertained on account of Theorem 5.

If applied to $g_0(x)$, next lemma suggests that the condition (26) might be established for a wide variety of cases, e.g., for the case where a constant penalty q (> 0) is imposed on the continuation of the control process.

LEMMA 2. *Let $\varphi(x)$ denote a finite solution of* (6), *and ρ a corresponding control law to φ. If*

$$\sum_{i=1}^{\infty} q(x_i, u_i) = \infty \qquad P_x{}^\rho\text{-a.s.} \quad \text{on} \quad \{\tau = \infty\} \tag{28}$$

for $x \in C$, then $P_x{}^\rho(\tau < \infty) = 1$ holds.

Proof. Since the sequence $z(n; \varphi)$ makes a martingale with respect to $P_x{}^\rho$, the following inequalities are justified:

$$\varphi(x) \geqslant E_x{}^\rho I_{\tau > n} \left\{ \sum_{i=1}^{n-1} q(x_i, u_i) + \varphi(x_n) \right\} \geqslant E_x{}^\rho I_{\tau=\infty} \sum_{i=1}^{n-1} q(x_i, u_i).$$

Suppose that $P_x{}^\rho(\tau = \infty) > 0$ for some $x \in C$. Then, as n increases, the last term approaches infinity due to (28). On the other hand, $\varphi(x)$ is assumed to be finite. Hence a contradiction is derived.

Remark. For the stochastic control problem in Section 3, let us consider a quadratic cost $q(x, u) = x'Rx + u'Qu$ as usual. Then, the following generous conditions meet (28) for any $\rho \in R$, i.e., the condition (26) holds:

1) $E\{w'f_2(x)' Rf_2(x) w\} \geqslant m$ holds on C for some $m > 0$, where primes denote the transpose.

2) The density function $\theta(w)$ is symmetric with respect to the origin.

Let us remark the inequality

$$\sum_{i=1}^{n+1} x_i'Rx_i \geqslant \sum_{i=1}^{n} w_i'f_2(x_i)' Rf_2(x_i) w_i H(w_i'f_2(x_i)' Rf_1{}^i),$$

where the scalar function $H(x)$ takes 1 or 0 according as $x \geqslant 0$ or $x < 0$, and $f_1{}^i = f_1(x_i, u_i)$. The condition (28) follows if the right-hand side tends to infinity w.p.l. This limit relation is implied by the fact that there exist positive constants a and b such that

$$W\{w \mid w'f_2(x)' Rf_2(x) wH(w'f_2(x)' Rf) \geqslant a\} \geqslant b$$

for every d_s-vector f and x. *Reductio ad absurdum* verifies this inequality on account of the conditions 1) and 2).

7. Optimal and Suboptimal Control Laws

On finding the optimum cost function, it becomes a great matter of concern to obtain an effective control law. Needless to say, the optimal control law does not always exist except for the case $\chi = 0$. However, the control law ρ that corresponds to $g_\chi(x)$ seems to be the likeliest candidate for optimal control. Next theorem presents a sufficient condition for ρ to be optimal.

THEOREM 6. *Let ρ denote the corresponding control law to $g_\chi(x)$. If*

$$\liminf_{n\to\infty} g_\chi(x_n) \geqslant \chi(x_1, x_2, ...) \qquad P_x{}^\rho\text{-a.s. on } \{\tau = \infty\}, \tag{29}$$

then ρ is the optimal control, i.e., $E_x{}^\rho z^\chi = g_\chi(x)$ holds.

Especially if for all x,

$$P_x{}^\rho(\tau < \infty) = 1 \tag{30}$$

then the conclusion is valid.

Proof. Noting that the sequence $z(n; g_\chi)$ makes a martingale with respect to $P_x{}^\rho$, we have

$$\begin{aligned} g_\chi(x) &\geqslant E_x{}^\rho \liminf_{n\to\infty} z(n; g_\chi) \\ &= E_x{}^\rho \left[\sum_{i=1}^{\tau-1} q(x_i, u_i) + I_{\tau<\infty} r(x_\tau) + I_{\tau=\infty} \liminf_{n\to\infty} g_\chi(x_n)\right]. \end{aligned}$$

on account of Fatou's lemma. Then, considering the inequality (29), we obtain $g_\chi(x) \geqslant E_x{}^\rho z^\chi$. The proof is completed.

Note that the condition (30) implies for the case $\chi = \infty$ that optimality of the control law ρ corresponding to $g_\infty(x)$ is equivalent to $\rho \in R^f$, i.e., each initial state $x \in C$ can be transferred w.p.l. to the terminal region T by taking the law ρ.

Further remarks are on the case $\chi = \infty$. Another approach for optimal control law is found in Kushner [10] that mainly discusses continuous time-parameter problem. With use made of Theorem 2, his method of comparison controls applies as follows. Let φ be a solution of (6). If the control law $\bar{\rho}$ corresponding to φ belongs to R^f, and if $I_{\tau>n}\varphi(x_n)$ are uniformly integrable under any $\rho \in R^f$, then $\bar{\rho}$ is optimal and $\varphi = g_\infty$.

Finally, a composite procedure for (η, ϵ)-optimal control law is described below. This method presents $g_\infty(x)$ constructively under the assumption that expectation of the terminating time is finite for every $\rho \in R_\infty$, although Theorem 3 needs preparatory contrivances in determination of the initial function.

THEOREM 7. *If*

$$E_x{}^\rho \tau < \infty \tag{31}$$

holds for every $\rho \in R_\infty$, *then the minimal solutions* $g^\delta(x)$ $(\delta > 0)$ *of the equations*

$$\begin{aligned} \varphi(x) &= r(x), & x \in T, \\ &= \delta + \min_{u \in U(x)} \left[q(x, u) + \int P(x, u; dy)\, \varphi(y) \right], & x \in C, \end{aligned}$$

converge nonincreasingly to $g_\infty(x)$. *Moreover, for any* $\epsilon > 0$, *there exists a* $\delta > 0$ *such that the control law corresponding to* $g^\delta(x)$ *is* (η, ϵ)*-optimal, if* $\int g^\delta(x)\, \eta(dx) < \infty$ *is satisfied for some* $\delta > 0$.

Proof. Let us consider another problem with the continuation cost $\delta + q(x, u)$, the terminal cost $r(x)$, and $\chi = \infty$. For descrimination, notations z_δ^∞ and R_∞^δ are used in this problem. Lemma 2 and Corollary in Section 6 show that $g^\delta(x)$ is the optimum cost function of the perturbed problem. Hence the inequality $g^\delta(x) \geqslant g_\infty(x)$ is immediate due to $z^\infty \leqslant z_\delta^\infty$. Since $g^\delta(x)$ decreases as δ tends to zero, its limit function $\tilde{g}(x)$ also satisfies (6), and $\tilde{g}(x) \geqslant g_\infty(x)$ follows.

It then suffices to show $\tilde{g} \in \Phi_\infty$ for the equality $\tilde{g}(x) = g_\infty(x)$. Note that $R_\infty \subset R_\infty^\delta$, i.e., $R_\infty = R_\infty^\delta$. In fact, the condition (31) implies

$$E_x^\rho z_\delta^\infty = E_x^\rho z^\infty + \delta E_x^\rho(\tau - 1) \tag{32}$$

for every $\rho \in R_\infty$. Then, on account of Theorem 2, the sequence $I_{\tau > n} g^\delta(x_n)$ is uniformly integrable under any $\rho \in R_\infty$. Recalling the inequality $g^\delta(x) \geqslant \tilde{g}(x)$, the required inclusion is obtained.

For the rest part of Theorem 7, let us consider (32) for the control law ρ^δ corresponding to $g^\delta(x)$. Then the inequalities

$$g^\delta(x) \geqslant E_x^{\rho^\delta} z^\infty \geqslant g_\infty(x)$$

follow directly from Theorem 4. Therefore, the monotone convergence theorem assures the result.

Appendix

Let $(X, \mathfrak{F})$ denote a σ-algebra, U a compact space with metric d and $K(U)$ the set of all nonempty closed subsets of U. A set-valued mapping Γ: $X \to K(U)$ is called measurable if

$$\Gamma^B \triangleq \{x \mid \Gamma(x) \cap B \neq \varnothing\} \in \mathfrak{F}$$

holds for every $B \in K(U)$. To begin with, auxiary lemmas are presented that are necessitated for the proof of the main result.

LEMMA 3. *Let a sequence of measurable mappings $\Gamma_j: X \to K(U)$ satisfy $\Gamma_j(x) \supset \Gamma_{j+1}(x)$, and let $\Gamma_\infty(x) \triangleq \bigcap_{j=1}^{\infty} \Gamma_j(x)$. Then, $\Gamma_\infty(x): X \to K(U)$ is measurable.*

Proof. By the assumption, $\{\Gamma_j^B\}$ is a nonincreasing sequence of measurable sets. In order to obtain the result, it is enough to show $\bigcap_{j=1}^{\infty} \Gamma_j^B \subset \Gamma_\infty^B$ considering $\Gamma_j^B \supset \Gamma_\infty^B$. For $x \in \bigcap_j \Gamma_j^B$, there exists u_j such that $u_j \in B$ and $u_j \in \Gamma_j(x)$. From the compactness of U, the sequence u_j has an accumulating point u_∞. Since B and $\Gamma_j(x)$ are closed subsets, it follows that $u_\infty \in B \cap \Gamma_j(x)$. Hence $u_\infty \in B \cap \Gamma_\infty(x)$, which implies $x \in \Gamma_\infty^B$.

LEMMA 4. *Let $\Gamma_3(x) \triangleq \Gamma_1(x) \cap \Gamma_2(x) \neq \varnothing$ for measurable mappings Γ_1 and Γ_2. Then $\Gamma_3: X \to K(U)$ is also measurable.*

Proof. Since U is separable, we can choose a countable dense subset $\{u_i\}$ of U. Let S_{ij} denote the closed ball of radius 2^{-j}, centered at u_i. Letting

$$\Gamma_k^j(x) = \bigcup_{\{i \mid \Gamma_k(x) \cap S_{ij} \neq \varnothing\}} S_{ij}$$

for $k = 1$ and 2, we obtain measurable mappings $\Gamma_k^j: X \to K(U)$ such that $\Gamma_k^j(x) \supset \Gamma_k^{j+1}(x) \supset \Gamma_k(x)$. Denoting $\Gamma_1^j(x) \cap \Gamma_2^j(x)$ by $\Gamma_3^j(x)$, the relation

$$\{x \mid \Gamma_3^j(x) \cap B \neq \varnothing\} = \bigcup_{\{i \mid S_{ij} \cap B \neq \varnothing\}} \bigcap_{k=1}^{2} \{x \mid \Gamma_k(x) \cap S_{ij} \neq \varnothing\}$$

holds for any $B \in K(U)$, which asserts the measurability of Γ_3^j.

By Lemma 3 and the relation $\Gamma_3^j(x) \supset \Gamma_3^{j+1}(x) \supset \Gamma_3(x)$, it suffices to show $\Gamma_3(x) \supset \bigcap_j \Gamma_3^j(x)$ in order to terminate the proof. Let $\bar{u} \in \bigcap_j \Gamma_3^j(x)$. From the definition of Γ_3^j, for each k and j, there exist $S_{i_1 j}$ and $S_{i_2 j}$ such that $\Gamma_k(x) \cap S_{i_k j} \neq \varnothing$ and $\bar{u} \in S_{i_k j}$. Then it follows immediately that $d(\bar{u}, \Gamma_k(x)) \leqslant 2^{-(j-1)}$ for all j. That is, the inclusion $\bar{u} \in \Gamma_k(x)$ is proved for $k = 1$ and 2.

LEMMA 5. *Let the function $f(x, u)$ be given by the limit of a nonincreasing sequence of $f_i(x, u)$ that are measurable in x and continuous in u, and $\Gamma: X \to K(U)$ be measurable. Then, the function $\varphi(x) \triangleq \min_{u \in \Gamma(x)} f(x, u)$ is measurable.*

Proof. If the conclusion is valid for $f = f_i$, then the limit relation

$$\lim_{i \to \infty} \min_{u \in \Gamma(x)} f_i(x, u) = \min_{u \in \Gamma(x)} f(x, u)$$

proves the required result. Let f be measurable in x and continuous in u. Noting that the function $\min_{u \in B} f(x, u)$ is measurable for any $B \in K(U)$, let us define a measurable function $\varphi^j(x) = \inf_i \varphi^{ij}(x)$, where

$$\varphi^{ij}(x) = \min_{u \in S_{ij}} f(x, u) \qquad \text{on} \qquad \{x \mid \Gamma(x) \cap S_{ij} \neq \varnothing\},$$

and ∞ elsewhere. Then $\varphi^j(x)$ comes to be the infimum of $f(x, u)$ in u, where u ranges over the intermediate term of

$$\Gamma(x) \subset \bigcup_{\{i \mid \Gamma(x) \cap S_{ij} \neq \varnothing\}} S_{ij} \subset \{u \mid d(u, \Gamma(x)) \leqslant 2^{-j}\} \triangleq \Gamma_j(x).$$

Thus,

$$\min_{u \in \Gamma_j(x)} f(x, u) \leqslant \varphi^j(x) \leqslant \varphi(x)$$

is obtained. Since the left-hand side tends to $\varphi(x)$ as $j \to \infty$, it follows that $\lim_{j\to\infty} \varphi^j(x) = \varphi(x)$, completing the proof.

Then our theorem of Filippov's type is stated as follows:

THEOREM 8. *Let the function $f(x, u)$ be such that presented in Lemma 5, and Γ: $X \to K(U)$ be measurable. Then, there exists a measurable function $u^0(x) \in \Gamma(x)$ that satisfies*

$$f(x, u^0(x)) = \min_{u \in \Gamma(x)} f(x, u).$$

Proof. By Lemma 5, the function

$$F(x, u) = f(x, u) - \min_{u \in \Gamma(x)} f(x, u)$$

also satisfies the condition of Lemma 5. Now letting $\bar{\Gamma}(x) = \{u \mid F(x, u) \leqslant 0\} \neq \varnothing$, we have the relation

$$\begin{aligned} \bar{\Gamma}^B &= \{x \mid F(x, u) \leqslant 0 \text{ for some } u \in B\} \\ &= \{x \mid \min_{u \in B} F(x, u) \leqslant 0\} \in \mathfrak{F} \end{aligned}$$

for any $B \in K(U)$. Thus the mapping $\bar{\Gamma}$: $X \to K(U)$ is measurable. Since the defininition of F obviously implies $\tilde{\Gamma}(x) \triangleq \bar{\Gamma}(x) \cap \Gamma(x) \neq \varnothing$, $\tilde{\Gamma}$ is a measurable mapping due to Lemma 4. For an arbitrary measurable mapping Γ: $X \to K(U)$, Kuratowski's selection theorem [11] assures the existence of measurable function $u^0(x) \in \Gamma(x)$. Therefore, noting the equality

$$\tilde{\Gamma}(x) = \{u \mid u \in \Gamma(x) \text{ and } f(x, u) \leqslant \min_{u \in \Gamma(x)} f(x, u)\},$$

the proof of Theorem 8 is terminated.

ACKNOWLEDGMENT

The author wishes to thank Dr. Y. Hattori, Professor of Kyoto University, for his helpful discussions and encouragements.

References

1. J. Neveu, "Mathematical Foundations of the Calculus of Probability," Holden–Day, San Francisco, 1965.
2. D. Blackwell, Discounted dynamic programming, *Ann. Math. Statist.* **36** (1965), 226–235.
3. R. E. Strauch, Negative dynamic programming, *Ann. Math. Statist.* **37** (1966), 871–890.
4. K. Hiderer, "Foundations of non-stationary dynamic programming with discrete time parameter," Lecture Notes in Operations Res. and Math. Sys., Vol. 33, Springer-Verlag, Berlin, 1970.
5. R. Bellman, "Dynamic Programming," Princeton Univ. Press, Princeton, N.J., 1957.
6. J. L. Doob, "Stochastic Processes," Wiley, New York, 1953.
7. A. Maitra, Discounted dynamic programming on compact metric spaces, *Sankhyā Ser. A.* **30** (1968), 211–216.
8. N. Furukawa, Markovian decision processes with compact action spaces, *Ann. Math. Statist.* **43** (1972), 1612–1622.
9. M. Schäl, On continuous dynamic programming with discrete time-parameter, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **21** (1972), 279–288.
10. H. J. Kushner, "Stochastic Stability and Control," Academic Press, New York, 1967.
11. K. Kuratowski and C. Ryll-Nardzewski, A general theorem on selectors, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **13** (1965), 397–403.